MR. YANN XAVIER CLAUDE BOURGEOIS (Orcid ID : 0000-0002-1809-387X)

DR. JORIS A. M. BERTRAND (Orcid ID : 0000-0002-3379-1019)

MR. BORJA MILÁ (Orcid ID : 0000-0002-6446-0079)

Article type : Original Article

Differential divergence in autosomes and sex chromosomes is associated with intra-island diversification at a very small spatial scale in a songbird lineage

Running title: Genomic divergence in an island songbird

Yann XC Bourgeois[1,2], Joris AM Bertrand[2,3], Boris Delahaie[2,4], Hélène Holota[2], Christophe Thébaud[2*], Borja Milá[5*]

[1] School of Biological Sciences, University of Portsmouth, Portsmouth PO1 2DY, UK

[2]Laboratoire Évolution et Diversité Biologique (EDB), UMR 5174 Centre National de la Recherche Scientifique (CNRS) - Université Paul Sabatier - Institut de Recherche pour le Développement (IRD), 118 Route de Narbonne, F-31062 Toulouse Cédex 9, France.

[3]Université de Perpignan Via Domitia, Laboratoire Génome & Développement des Plantes, UMR 5096, 58

av. Paul Alduy, Bât. T, F-66860 Perpignan Cédex 9, France.

[4]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom.

[5]National Museum of Natural Sciences, Spanish National Research Council (CSIC), Madrid 28006, Spain.

Corresponding author: Yann Bourgeois, E-mail: yann.x.c.bourgeois@gmail.com

*These authors contributed equally.

Abstract

Recently diverged taxa showing marked phenotypic and ecological diversity are optimal systems to understand the genetic processes underlying speciation. We used genome-wide markers to investigate the diversification of the Reunion grey white eye (*Zosterops borbonicus*) on the small volcanic island of Reunion (Mascarene archipelago), where this species complex exhibits four geographic forms that are parapatrically distributed across the island and differ strikingly in plumage colour. One form restricted to the highlands is separated by a steep ecological gradient from three distinct lowland forms which meet at narrow hybrid zones that are not associated with environmental variables. Analyses of genomic variation based on SNP data from genotyping-by-sequencing and pooled RADseq approaches, reveal that signatures of selection associated with elevation can be found at multiple regions across the genome, whereas most loci associated with the lowland forms are located on the Z sex chromosome. We identified *TYRP1*, a Z-linked colour gene, as a likely candidate locus underlying colour variation among lowland forms. Tests of demographic models revealed that

highland and lowland forms diverged in the presence of gene flow, and divergence has progressed as gene flow was restricted by selection at loci across the genome. This system is promising to investigate how adaptation and reproductive isolation shape the genomic landscape of divergence at multiple stages of the speciation process.

Keywords: Speciation, Sex chromosome, Natural selection, Genomics, Population Differentiation, Plumage colour evolution, *Zosterops*.

Introduction

As populations and lineages diverge from each other, a progressive loss of shared polymorphisms and accumulation of fixed alleles is expected. This is influenced by neutral processes (e.g. genetic drift), but also natural and sexual selection, and the interaction between these processes may vary between different parts of the genome, creating a mosaic pattern of regions displaying different rates of divergence (Wu, 2001; Nosil et al., 2009). However, genomic regions directly involved in local adaptation and reproductive isolation may experience reduced effective gene flow compared to the genomic background (Ravinet et al., 2017). In addition, the effects of selection at linked sites can also locally increase divergence and magnify the effects of non-equilibrium demography over large genomic regions (Cruickshank and Hahn, 2014; Burri et al., 2015; Burri, 2017; Van Belleghem et al., 2018). Thus, establishing how different processes such as drift, selection and also gene flow, shape

the rates of divergence at the genomic scale is critical to understand the links between speciation processes and their genetic and genomic consequences (Gavrilets, 2014).

Identifying the main drivers of genome-wide differentiation (i.e. isolation by environment vs. reproductive isolation driven by non-adaptive factors) remains a complex question (Cruickshank and Hahn, 2014; Wolf and Ellegren, 2016; Ravinet et al., 2017). Recent studies have displayed varied results, and those focusing on the early stages of speciation have often emphasised ecological divergence over sexual selection and intrinsic incompatibilities (Bierne et al., 2011; Seehausen et al., 2014). In this context, studies of closely related taxa or populations that show phenotypic and ecological diversity, and are at different stages of divergence, hold promise to help clarify the chronology and relative importance of these underlying evolutionary mechanisms (Sætre and Sæther, 2010; Pryke, 2010; Safran et al., 2013; Seehausen et al., 2014; Delmore et al., 2015; Mořkovský et al., 2018).

We used the Reunion grey white-eye (*Zosterops borbonicus;* taxonomy following Gill and Donsker 2019), a songbird endemic to the small (2512 km^2) volcanic ocean island of Reunion (Mascarene archipelago, southwestern Indian Ocean), to quantify genome-wide patterns of divergence across its range and better understand underlying evolutionary factors. This species is characterized by complex patterns of plumage colour and size variation with five distinct variants recognised across the island (Gill 1973). These variants can be grouped into four parapatrically distributed geographical forms with abutting ranges that came into secondary contact after diverging in allopatry (Cornuault et al., 2015; Bertrand et al., 2016; Delahaie et al., 2017). Three lowland forms differ primarily in plumage colour (Gill, 1973; Cornuault et al., 2015) and show a unique distribution pattern, with each form being separated from the other two by narrow physical barriers such as rivers or lava fields (Gill, 1973). These forms differ strikingly in head coloration and include a light brown form (lowland brown-headed brown form; hereafter LBHB), a grey-headed brown form (GHB) with a brown back and a grey head, and a brown-naped brown form (BNB) with a brown back and nape, and a grey crown (Figure 1 and Figure S1; see Cornuault et al., 2015 for a detailed description). A fourth form, restricted to the highlands (between 1,400 m and 3,000 m), is relatively larger than the lowland forms and comprises two very distinct colour morphs, with birds showing predominantly grey (GRY) or brown (highland brown-headed brown form, HBHB) plumage, respectively (Gill, 1973; Milá et al.,

2010; Cornuault et al., 2015; Bertrand et al., 2016). Both morphs occur in sympatry and represent a clear case of plumage colour polymorphism (Bourgeois et al. 2017). This highland form is separated from all three lowland forms by relatively narrow contact zones located along the elevational gradient (Gill, 1973). One such contact zone was recently studied and was found to correspond to an ecotone between native habitat (> 1400 m above sea level; a.s.l.) and anthropogenic landscapes (< 1400 m a.s.l.), suggesting a possible role of environmental differences in influencing the location of these zones (Bertrand et al. 2016). While plumage colour differences between the two apparently similar all-brown variants (LBHB and HBHB) may appear subtle, they are in fact significant when considering bird vision and using a visual model to project these colours in an avian appropriate, tetrachromatic colour space (Cornuault et al. 2015). Patterns of colouration among forms and morphs are stable over time, with no apparent sex effect (see Gill 1973, Milá et al. 2010).

Recent studies have revealed that dispersal and gene flow must be limited in the Reunion grey white-eye, with low levels of historical and/or contemporary gene flow among populations, unless very close geographically (< 10 km) (Bertrand et al., 2014), and that more variation exists among the different geographical forms than would be expected under drift for both morphological and plumage colour traits (Cornuault et al. 2015). Thus, it is the combination of reduced dispersal and divergent selection that seems to explain why white-eyes were able to differentiate into multiple geographical forms within Reunion, as originally proposed by Frank Gill (Gill, 1973). The island exhibits a dramatic topography (maximum elevation : 3070 m) and a steep elevational gradient was found to be associated with strong divergent selection on phenotypes and marked genetic structure for autosomal microsatellites, a pattern that is consistent with isolation by ecology between lowland and highland forms (Bertrand et al., 2016). In contrast, lowland forms show no association with neutral genetic (microsatellite) structure or major changes in vegetation characteristics and associated climatic variables and are separated by very narrow hybrid zones centered on physical barriers to gene flow (Delahaie et al., 2017). Although the autosomal markers used could not provide information on sex-linked loci, these patterns of genetic differentiation suggest that while a sharp ecological transition between lowlands and highlands could drive differentiation at many autosomal loci through local adaptation, phenotypic divergence between lowland forms involves either fewer loci, or loci concentrated in a narrower genomic region not covered by microsatellites (Delahaie et al. 2017).

In this work, we aim at (i) identifying the genomic variation associated with phenotypic differentiation between forms in relation to ecological variation (low vs. high elevation) and divergence in signaling traits (conspicuous variation in plumage colour between lowland forms in the absence of abrupt ecological transitions), (ii) determining whether divergence peaks are found on autosomes or sex chromosomes, respectively, and (iii) identifying potential candidate genes associated with divergent genomic regions. We used individual genotyping by sequencing (GBS) (Elshire et al., 2011) to characterise the amount of divergence between forms. We further used a pooled RAD-sequencing (Baird et al., 2008) approach that produced a high density of markers to characterise with greater precision the genomic landscape of divergence and assess the extent of differentiation between the different colour forms. Finally, in order to test whether incomplete lineage sorting or gene flow explained shared genetic variation among forms, we use coalescent models to test alternative demographic scenarios of divergence, including models with different temporal patterns of gene flow and effective population size changes over time.

Material and Methods

Field sampling

We sampled a total of 259 Reunion grey white-eyes between 2007 and 2012 from nine locations that were chosen to extensively cover the species' range and the different geographic forms. We also sampled 25 Mauritius grey white-eyes (*Zosterops mauritianus*) from a single location on Mauritius to be used as an outgroup in some of our analyses since this species and the Reunion grey white-eye are sister taxa (Warren et al. 2006). Birds were captured using mistnets, marked with a uniquely numbered aluminum ring, and approximately 50 µL of blood were collected from each individual and preserved in Queen's lysis buffer (Seutin et al., 1991). All manipulations were conducted under a research permit issued by the Centre de Recherches sur la Biologie des Populations d'Oiseaux

(CRBPO) – Muséum National d'Histoire Naturelle (Paris). Individuals were sexed using PCR (Griffiths et al., 1998) in order to infer the number of distinct Z chromosomes included in each genetic pool. We included 152 females and 132 males in this study, among which 47 females and 48 males were included in the genotyping-by-sequencing experiment (see below).

Genotyping by sequencing (GBS) using individual DNA samples

We performed genotyping by sequencing (GBS hereafter; (Elshire et al., 2011) on 95 individuals, including 90 Reunion and five Mauritius grey white-eyes (Figure 1). We included 7-14 individuals from each Reunion location and two locations per geographic form; such a sampling scheme should be sufficient to retrieve patterns of differentiation and diversity at the scale of forms, as highlighted by both theoretical (Willing et al., 2012) and empirical studies (Jeffries et al., 2016; Nazareno et al., 2017). GBS is similar to RAD-sequencing, but involves fewer preparation steps (Elshire et al., 2011) and samples loci at a lower resolution across the genome. Approximately one microgram of DNA was extracted with a QIAGEN DNeasy Blood & Tissue kit following manufacturer's instructions and sent to the BRC Genomic Diversity Facility at Cornell University (see Elshire et al., 2011) for single-end sequencing on a single lane of an Illumina HiSeq2000 after digestion with *PstI*. Read length was 100 bp. Three individuals had to be removed from subsequent analyses due to the extremely low number of reads obtained (Supplementary Table 1). Raw reads were trimmed with Trimmomatic (v 0.33, (Bolger et al., 2014)) with a minimum base quality of 20. We used the recently assembled *Z. lateralis* genome (Cornetti et al., 2015) to map reads back onto this reference with BWA MEM (v. 0.7.12, (Li and Durbin, 2009)) and SAMTOOLS (v 1.3.1, (Li et al., 2009)), instead of creating consensuses directly from data as in Bourgeois et al. (2013). Reads with a mapping score below 20 were excluded (samtools view -q 20). We then aligned contigs and scaffolds from a congeneric white-eye species *Zosterops lateralis* on the Zebra Finch (*Taeniopygia guttata*) passerine reference genome (version July 2008, assembly WUGSC v.3.2.4) using LASTZ (v 1.03.54, (Schwartz et al., 2003; Harris, 2007)). We used the following options and thresholds:--masking=254 --hspthresh=4500 --gappedthresh=3000. The first option means that any locus found mapping more than 254 times is automatically masked and does not appear in the final pairwise alignment. The –hspthresh parameter is an option that excludes any alignment with a score lower than 4500 during the gap-free extension

stage. The --gappedthresh option controls the maximum size of the gaps allowed to join best local alignments; the higher the score, the fewer gaps are allowed. We used the same set of options previously used in comparisons between other related bird genomes, such as chicken and grouse (e.g. (Kozma et al., 2016)). Scaffolds were then assigned to chromosomal regions based on their alignment scores. We note that synteny is well conserved in birds (Derjusheva et al., 2004), and that misalignment is therefore unlikely to constitute a major source of errors.

SNPs were called using freebayes (v 0.9.15-1, (Garrison and Marth, 2012)) and filtered with VCFTOOLS (v 0.1.12b) using the following criteria for autosomal markers: (i) a sequencing depth between 8 and 100X for each individual genotype; (ii) a minimal genotype quality of 20; and (iii) no more than 9 missing genotypes. Missing data per individual before filtering and after removing individuals with low read count was at most 50% (average 30%, s.d.=5%, see Suppl. Table 1). Average sequencing depth was 9.75X (s.d.=3.05). The range of sequencing depth for filtering was chosen based on visual examination of histograms produced by SAMTOOLS (option depth), to remove loci with a clear excess of mapping reads that may indicate repetitive sequences and those loci with very low depth for which genotypes may not be called confidently, while retaining enough information for inference. For Z-linked markers, we first listed scaffolds mapping on the Zebra finch's Z chromosome based on LASTZ alignments. We then used VCFTOOLs to extract genotypes found on Z scaffolds (providing a list of these scaffolds with the option –bed). SNPs were filtered in male individuals only, using the same criteria as for autosomes, except that no more than 5 missing genotypes were allowed. We then extracted these sites in females only, using VCFTOOLS (option --positions), and removed markers displaying more than three heterozygous females, allowing for some tolerance since freebayes attempts to balance the count of heterozygotes in a diploid population.. This led to the removal of 171 sites out of 1,136. We then recalled SNPs with freebayes in females only, assuming haploidy (option –ploidy 1). The constraint on sequencing depth and genotype quality was removed in females to consider the fact that a single Z copy is found in these individuals, therefore reducing depth of coverage at Z-linked markers. Since we excluded reads with a mapping quality below 20 when creating BAM files, we considered that a single read was enough to call a site in females for the Z scaffold. This decision was taken to maximize the number of markers available for this chromosome. The final dataset consisted of 34,951 autosomal markers and 965 Z-linked markers.

Recent studies have suggested the existence of a neo sex chromosome in Sylvioidea, consisting of a fusion between ancestral Z and W chromosomes with the first 10 Mb of the Zebra Finch's chromosome 4A (Pala et al., 2012). We therefore excluded this region from our analyses and studied it separately, focusing on males only. In males, 917 and 730 SNPs called by freebayes were found polymorphic on the Z and the 4A sex-linked fragment, respectively.

Pooled RAD-seq

To identify loci and genomic regions associated with ecological variation (low vs. high elevation) and divergence plumage colour between lowland forms, we used a paired-end RAD-sequencing protocol, using a dataset partially described elsewhere (Bourgeois et al., 2013) in which six pools of 20-25 individuals from the same three locations as those sampled for the high-elevation form in the GBS experiment were sequenced. We added seven more pools of 16-25 individuals from the lowland forms to cover the same localities than the GBS dataset **(**Figure 1, Suppl. Table 2). This protocol was used since it produced a higher density of markers along the genome relative to the GBS approach described above, thus increasing the ability to detect outlier genomic regions. This approach resulted in approximately 600,000 contigs with an average size of 400 bp, covering about 20% of the genome (Bourgeois et al., 2013). The larger number of individuals included in each pool should also increase the ability to detect shifts in allele frequencies between populations. We modified the bioinformatics protocol used in Bourgeois et al. (2013) by mapping the reads on the *Z. lateralis* genome using BWA MEM instead of creating contigs from the RAD-seq reads. PCR duplicates were removed using SAMTOOLS (Li et al., 2009). SNPs were called using Popoolation2 (v 1.201, (Kofler et al., 2011)), using a minimal allele count of two across all pools, and a depth between 10X and 300X for each pool to remove loci that were clear outliers for sequencing depth while keeping a high density of markers along the genome. We used BEDTOOLS (v 2.25.0, (Quinlan and Hall, 2010)) to estimate the proportion of sites covered at a depth between 10X and 300X in each pool (option genomecov). Overall, more than 1,104,000 SNPs for autosomes and 42,607 SNPs for the Z chromosome were obtained, covering between 12 and 18% of the genome (Sup. Table 2). We accounted for the unequal number of alleles between autosomal markers and Z-linked markers in all subsequent analyses.

Genetic structure

To assess population genetic structure within and between geographic forms, we first performed a principal components analysis (PCA (Patterson et al., 2006)) on all GBS autosomal markers, using the Bioconductor package SeqVarTools (v 1.24.0, (Huber et al., 2015)), excluding markers with a minimal allele frequency below 0.05. We then evaluated population structure for both autosomal and Z-linked markers using the software ADMIXTURE (v 1.3.0, (Alexander and Novembre, 2009)). This software is a fast and efficient tool for estimating individual ancestry coefficients. It does not require any a-priori grouping of individuals by locality but requires defining the expected number (K) of clusters to which individuals can be assigned. Importantly, ADMIXTURE allows specifying which scaffolds belong to sex chromosomes, and corrects for heterogamy between males and females. "Best" values for K were assessed using a cross-validation (CV) procedure using 10 CV replicates. In this context, cross-validation consists in masking alternatively one fifth of the dataset, then using the remaining dataset to predict the masked genotypes. Predictions are then compared with actual observations to infer prediction errors. This procedure is therefore sensitive to heterogeneity in structure across markers induced by, e.g., selection. Therefore, we present results for all values of K as they may reveal subtle structure supported by only a subset of markers under selection. Based on patterns of LD-decay, we thinned the dataset to limit the effects of linkage, with a minimal distance between two adjacent markers of 1,000 bp (Figure S2). Pairwise LD (measured as r^2, which does not require phasing) between all pairs of markers was computed in VCFTOOLS (v 0.1.12b).

To further explore whether changes in SNP caller and the number of markers could affect the ADMIXTURE analysis and observed differences between autosomes and Z-linked markers, we called Z-linked SNPs using ANGSD (v 0.923, (Korneliussen et al., 2014)), following an approach similar to the one used with freebayes. We first called SNPs in all individuals assuming diploidy, using a uniform prior based on allele frequencies but not assuming Hardy-Weinberg equilibrium in samples (option -doPost 2). SNP likelihoods were computed following the model implemented in SAMTOOLS (-GL=1). We then called SNPs in females only, assuming haploid markers and calling the consensus base (option -doHaploCall 2). We filtered reads so they mapped to a single site in the genome (-uniqueOnly 1 -remove_bads 1), had a mapping quality of at least 20 (-minMapQ 20), a minimum read quality of 20 (-minQ 20), and were covered in at least two thirds of individuals with a

minimum individual depth of 6X in males (-geno_minDepth 6). We corrected for excessive mismatch with the reference and excess of SNPs with indels (-C 50 -baq 1). This resulted in 2,282 Z-linked SNPs.

To assess the relative proportion of genetic variance contributing to the differentiation of geographic forms (estimated by F_{CT}) while taking into account population substructure (F_{SC} and F_{ST}), we conducted an analysis of molecular variance (AMOVA) in Arlequin v3.5 (Excoffier and Lischer, 2010) using as groups either islands (Reunion vs. Mauritius), lowlands and highlands, or the forms themselves. *F*-statistics for the whole dataset are weighted averages. Significance was assessed with 1000 permutations.

We assessed relationships between populations from pooled data using POPTREE2 (Takezaki et al., 2010) to compute F_{ST} matrices across populations using allele frequencies. A Neighbor-Joining tree was then estimated from these matrices. We included 20,000 random SNPs with a minimum minor allele count of 2. Branch support was estimated through 1,000 bootstraps. As a supplementary control, we also report the correlation matrix estimated from the variance-covariance matrix obtained by the software BAYPASS (v2.1) (Gautier, 2015) for both GBS and pooled data. The variance-covariance matrix reflects covariation of allele frequencies within and between populations. The correlation matrix describing pairwise relatedness between populations was then derived using the R function cov2cor() provided with BAYPASS. The function hierclust() was used to perform hierarchical clustering based on matrix coefficients. For POPTREE2 analysis and AMOVAs, we took care of accounting for the different number of alleles between males and females for Z-linked markers, including one allele for females and two for males.

Demography

To help distinguish between the presence of extensive gene flow between forms and incomplete lineage sorting as an explanation for the generally low differentiation, we performed model comparison under the likelihood framework developed in fastsimcoal2.6 (Excoffier et al., 2013) using frequency spectra inferred by ANGSD from autosomal and Z-linked GBS data. We focused on the split between populations from high and low elevations. This split was the clearest across all analyses,

with very little genetic differentiation between lowland forms for these markers (see Results). Given the weak genetic substructure for autosomal markers, and since a large number of individuals within groups is required to infer very recent demographic events (Robinson et al., 2014), we pooled individuals into three groups: highlands, lowlands and Mauritius. We acknowledge that models incorporating substructure could be built, but this would come at the cost of adding more parameters. We preferred to limit this study to simple models that can serve as a basis for future, more detailed work using more markers and individuals (Otto and Day, 2007).

We used GBS autosomal and Z-linked markers as they could be filtered with higher stringency than pooled markers and were more likely to follow neutral expectations, with only 1.6% of genetic variance explained by GBS loci harboring significant differentiation between highland and lowland forms (see AMOVAs in Results). SNPs mapping on scaffolds corresponding to the neo sex chromosome region on 4A were discarded from the analysis (see Results). We extracted the joint derived site frequency spectrum (SFS) using ANGSD, which takes into account genotypic uncertainties to directly output the most likely SFS. We used the reference genome as an outgroup to assign alleles to ancestral or derived states. Using ANGSD should correct for biases that may happen when calling genotypes from low and medium-depth sequencing. We filtered markers using the same criteria used for the ADMIXTURE analysis on Z-linked loci (-uniqueOnly 1 -remove_bads 1 -minMapQ 20 -minQ 20 -doPost 2 -geno_minDepth 6 -C 50 -baq 1), but excluded two individuals with very low average sequencing depth (993 and 11_0990), and removed sites that were not covered in all remaining individuals. For Z-linked markers, we extracted the SFS from male individuals only, given that ANGSD cannot simultaneously extract the spectrum from samples with mixed ploidy. Entries in the joint site frequency spectrum were examined to exclude potential paralogs displaying strong heterozygosity.. We did not filter on minimal allele frequency since singletons are important to properly estimate parameters and likelihoods. We compared four distinct demographic models (Figure 3), one in which all forms were essentially treated as a single population (by forcing a very recent split 10 generations ago) that went through a change in effective population sizes after the split from Mauritius, one with no gene flow between highland and lowland forms, one allowing constant and asymmetric gene flow between lowland and highland forms, and a last model where gene flow could vary at some time in the past between present time and the split between lowland and highland forms.

Population sizes could vary at each splitting time and each group was assigned a specific effective population size. Parameters were estimated from the joint SFS using the likelihood approach implemented in fastsimcoal2.6 (Excoffier et al., 2013). Parameters with the highest likelihood were obtained after 20 cycles of the algorithm, starting with 50,000 coalescent simulations per cycle, and ending with 100,000 simulations. This procedure was replicated 50 times and the set of parameters with the highest final likelihood was retained as the best point estimate. The likelihood estimated by fastsimcoal2.6 is a composite likelihood, which can be biased by covariance between close markers (Excoffier et al., 2013). To properly compare likelihoods and keep the effects of linkage to a minimum, we first used a thinned dataset with SNPs separated by at least 10,000 bp (see LD decay, Figure S2). We then used the complete dataset for parameter estimation. We used a fixed divergence time of 430,000 years between Reunion and Mauritius grey white-eyes (Warren et al., 2006) and assumed a generation time of one year to calibrate parameters and obtain an estimate in demographic units for the timing of diversification in the Reunion grey white eye. We used python scripts implemented in ∂a∂i (Gutenkunst et al., 2009) to visualise and compare predicted and observed SFS. To assess deviation from neutrality and stable demography, we estimated Tajima's *D* (Tajima, 1989) from the spectra with ∂a∂i.

We estimated 95% confidence intervals (CIs) using a non-parametric bootstrap procedure, bootstrapping the observed SFS 100 times using ANGSD and repeating the parameter estimation procedure on these datasets, using 10 replicates per bootstrap run to reduce computation time. To visualise whether the model fitted the observed data, we compared the observed SFS with 100 SFS simulated using the set of parameters obtained from the model with the highest likelihood. Coalescent simulations were carried out in fastsimcoal2.6, simulating DNA fragments of the size of GBS loci (91bp) and using a mutation rate of 3.6×10^{-9} mutations/generation (Axelsson et al., 2004) until the number of SNPs matched the number of segregating sites in the observed dataset. Parameters for each simulation were uniformly drawn from the CIs of the best model. We summarized SFS by PCA using the gfitpca() function of the abc package (Csilléry et al., 2012), including only entries with at least 0.1% of the total number of segregating sites to reduce variance.

Selection and environmental association

To detect loci displaying a significant association with geographic forms and environment, we performed an association analysis on the pooled RAD-seq data. We used the approach implemented in BAYPASS (v 2.1.)) to detect SNPs displaying high differentiation (Gautier, 2015). This approach is designed to robustly handle uncertainties in allele frequencies due to pooling and uneven depth of coverage, by directly using read count data. It performs well in estimating differentiation and population structure (Hivert et al., 2018). Briefly, BAYPASS estimates a variance-covariance matrix reflecting correlations between allele frequencies across populations. Divergence at each locus is quantified through a Bayesian framework using the X^TX statistic, which can be seen as a SNP-specific F_{ST} corrected by this matrix. BAYPASS also offers the option to estimate an empirical Bayesian *P*-value (*eBPis*) which can be seen as the support for a non-random association between alleles and any population-specific covariable. We also computed *eBPis* to determine the level of association of each SNP with geographic form and elevation. Specifically, we tested associations between allele frequencies and a binary covariable stating whether pools belonged or not to Mauritius grey white-eye, GHB, BNB and LBHB forms. We also tested association with elevation, coded as a continuous variable. BAYPASS was run using default parameters under the core model with 25 pilot runs and a final run with 100,000 iterations thinned every 100 iterations. Because estimates of the variance-covariance matrix are robust to minor allele count thresholds (Gautier, 2015), we only included SNPs with a minor allele count of 10 over the entire dataset to reduce computation time and ran separate analyses on Z chromosome and autosomes to account for their distinct patterns of differentiation and allele counts.

GO enrichment analysis

To gain insight into the putative selective pressures acting on Reunion grey white-eye populations, we performed a Gene Ontology (GO) enrichment analysis, selecting for each association test SNPs in the top 1% for both *eBPis* and X^TX, considering separately the Z chromosome and autosomes. Gene annotations within 100 kb windows flanking selected SNPs were extracted using the zebra finch reference and the intersect function in bedtools v2.25.0 (Quinlan and Hall, 2010). We chose 100kb based on the average territory size of protein-coding genes in the zebra finch (87.8kb, see Table 1 in

Warren et al., 2010). We adjusted the gene universe by removing zebra finch genes not mapping onto the *Z. lateralis* genome. GO enrichment analysis was performed using the package topGO in R (Alexa and Rahnenfuhrer, 2016), testing for significant enrichment using a Fisher's test for overrepresentation. We present raw *P*-values instead of *P*-values corrected for multiple testing, following recommendations from the topGO manual, and because we wanted to detect any interesting trend in the dataset that could then be further explored. We present the top 50 GO terms associated with biological processes, ranked by their raw *P*-values.

Results

Genetic structure and relationships among geographic forms

We first assessed whether geographic forms could be distinguished based on the genomic data available for individuals (GBS). A principal components analysis on autosomal allele frequencies revealed a clear distinction between Mauritius and Reunion grey white-eyes on the first axis, as well as a distinction between localities from lowlands and highlands on the second axis (Figure 1B). When excluding Mauritius white-eyes, the main distinction remained between localities from high and low elevation, reflecting differentiation between highland and lowland forms. Further principal components did not reveal any strong clustering of the different forms from the lowlands (Figure 1B). This pattern was further confirmed by the ADMIXTURE analysis, where low and high clustered separately for both autosomal and Z-linked markers (Figure 2). The cross-validation procedure gave K=2 and K=7 as best models for autosomal and Z-linked markers respectively (Figure S3). Cross-validation scores were however low for all K values ranging from 5 to 7 for Z-linked markers, making it difficult to clearly identify an optimal value. Given the subtle genetic structure, we present results for values of K ranging from 2 to 7.

Clustering was consistent with the PCA, with a distinction between localities from high and low elevations at K=2 and 3, for Z-linked and autosomal markers respectively (Figure 2A). For autosomal markers, higher values of K highlighted a structure consistent with sampling sites, but structure

according to geographic forms was more elusive. For Z-markers, clustering tended to be more consistent with forms at low elevation. At *K*=7, one cluster corresponded to localities at high elevation, three others grouped lowland localities by forms, while a fifth cluster included Mauritian individuals (Figure 2A). Two localities (LF and TC) displayed stronger signals of mixed ancestry, probably due to gene flow between those two localities which are close to a zone of contact between the GHB and BNB forms. The same pattern was observed using the set of Z-linked SNPs called with ANGSD, the distinction between GHB and BNB forms being even clearer, likely due to the larger number of markers (Figure S4). We note that using this set of markers improved the discrimination between colour forms, probably because of the higher number of SNPs remaining after filtering with ANGSD (2,282 loci) instead of freebayes (965 loci).

Recent studies revealed the existence of neo sex chromosomes in Sylvioidea, to which Zosteropidae belong (Pala et al., 2012), and their existence was recently confirmed in the Reunion grey white-eye (Leroy et al. 2019). These neo-sex chromosomes emerged from the fusion of Z and W chromosomes with the first 10 Mb of chromosome 4A. We extracted GBS markers mapping to this region and confirmed that they were sex-linked, as shown by a PCA on allele frequencies (Figure S5). Visual examination of genotypes revealed that females (ZW) displayed a strong excess of heterozygous markers, contrary to males (ZZ), due to divergence between neo-Z and neo-W chromosomes. We further investigated population structure in males only, using a set of combined Z and 4A markers, and an ADMIXTURE analysis tended to better discriminate geographic forms at *K*=6 with this set of markers (Figure S6).

The same pattern was observed with the POPTREE2 analysis on the pooled dataset. The autosomal topology supported a grouping of localities from high elevation together and supported a grouping of localities belonging to BNB and LBHB forms, yet there was no support for grouping together localities from the GHB form. In contrast, topology based on Z markers provided good support for a grouping of localities by geographic form (Figure 2B).

We further investigated population structure by estimating the variance-covariance matrix obtained from allele frequencies for both pooled and GBS data using BAYPASS (Gautier, 2015). Localities from high elevation were systematically found clustering together, a pattern consistent with previous analyses (Figure S7). Again, both analyses on Z-linked markers for GBS and pooled data revealed a

closer relationship between populations from the same geographic form when compared to autosomal markers. LBHB and BNB forms clustered together, with the GHB form branching off first within the lowland group.

To estimate quantitatively the proportion of the genome discriminating among geographic forms while accounting for population structure within forms, we performed a molecular analysis of variance (AMOVA) on GBS data. This analysis confirmed the previous patterns, with a proportion of variance explained by forms or elevation not higher than 1.6% for autosomal markers (Table 1). The strongest differentiation was observed between localities from low and high elevations and between Mauritius and Reunion populations. For Z-linked markers however, the proportion of variance explained by forms and elevation was more substantial, ranging from 3.4 to 12.6% (Table 1). This higher differentiation could not be explained by differences in sample size between autosomal and Z-linked data. For autosomal markers, AMOVAs performed only on males did not show substantial deviations from the results obtained with the full dataset (Table 1).

Demographic history

We compared four nested models (Figure 3), allowing for no or constant gene flow after the split between populations. The highest likelihoods were found for the model allowing change in gene flow at some time (T_{change})after the initial split at T_{split} between lowlands and highlands (model D, Figure 3), while the other models were clearly rejected (Table 2). Such a pattern of higher gene flow in recent times is consistent with a scenario of introgression through secondary contact.

Assuming a conservative divergence time between *Z. borbonicus* and *Z. mauritianus* of 430,000 years, parameter estimates suggested a split between high and low elevation populations 400,000 years ago and an increase in gene flow 80,000 years ago (Table 3). Overall, point estimates of effective migration rates ($2Nm$, with N the diploid population size) were high ($2N_{High}\, m_{Low\rightarrow High}$ =0.1 gene copies/generation, $2N_{Low}\, m_{High\rightarrow Low}$ =9.2 before T_{change}, then reaching $2N_{High}\, m_{Low\rightarrow High}$ = 9.9 and $2N_{Low}\, m_{High\rightarrow Low}$ = 6.0 80,000 years ago), consistent with homogenisation of genomes through migration. This model was able to explain our observed dataset, as indicated by visual comparisons of

the observed and predicted SFS (Figure S8A). Coalescent simulations using parameters drawn from CIs of the best model produced SFS similar to the observed one (Figure S8B).

Genes that are involved in reproductive isolation between populations are expected to resist gene flow due to counterselection of maladapted or incompatible alleles. This should result in an increased divergence at these loci when compared to the genomic background (Cruickshank and Hahn, 2014). We tested whether reduced gene flow explained increased divergence in the Z by estimating parameters from the four models but using this time all Z-linked markers, because they explained most of the differentiation between forms. We used the same filtering criteria that we used for autosomes but focused our analysis on diploid males only. The highest likelihood was also found for the model allowing gene flow to change after T_{split} (Table 2), again providing support for a scenario of introgression trhough secondary contact. As expected, given hemizygosity, effective population size estimates were lower for this set of markers than for autosomes (Table 3). Estimates of gene flow were also historically lower than for autosomes, while time since the split between highland and lowland groups was similar. This combination of lower population sizes and lower gene flow is expected to lead to increased divergence at Z-linked loci, in accordance with the stronger differentiation observed at these markers in AMOVAs and other tests.

To further explore whether differences in demographic inferences may be due to stronger effects of linked selection on the Z chromosome, we computed Tajima's *D* for each group (Tajima, 1989). This statistic should be negative (<2) in the case of recent population expansion, or if positive/purifying selection is acting. It should be positive (>2) in the case of a recent bottleneck or balancing selection. Tajima's *D* values for the autosomal spectra obtained by ANGSD were -0.29, -0.73 and -1.18 for Mauritius, highlands and lowlands respectively. Tajima's *D* values were higher for Z-linked markers, at 0.12, 0.09 and -0.76, but followed the same trend, being lower in lowlands than in highlands and Mauritius. This suggests that widespread effects of linked selection on the Z chromosome are unlikely to explain its higher differentiation.

Genome scan for association and selection analysis

We used BAYPASS (Gautier, 2015) on pooled data to retrieve markers displaying high levels of differentiation (X^TX) and association (empirical Bayesian *P*-values; *eBPis*) with 5 different features (elevation, being GHB, being BNB, being LBHB, being Mauritian). Overall, this pooled dataset included a total of 284 individuals sequenced at an average depth of ~250X over all pools. We examined Z and autosomal markers separately due to their distinct demographic histories. Results revealed a strong association of Z-linked markers with geographic forms and their associated colour phenotypes and elevation (Figure 4). SNPs discriminating Mauritius from other populations were found distributed uniformly along the genome. Most of the peaks displaying a large X^TX were also found associated with elevation. Strikingly, the clearest peaks on chromosomes 2, 3 and 5 covered large genomic regions, spanning several hundreds of kilobases. The sex-linked region on chromosome 4A was the clearest outlier. Since the neo-W chromosome is highly divergent and found in all females, an excess of variants with frequencies correlated to the proportion of females in the pool is expected. This may lead to high differentiation between pools with different sex ratios. Despite this, the strong association with elevation and geographic form on chromosome 4A is genuine since the expected proportion of divergent W-linked alleles in each pool was not correlated with those variables in our experiment (Figure S9), making confounding effects unlikely.

To assess the possible role of these genomic regions in adaptation, we performed a Gene Ontology (GO) enrichment analysis using the zebra finch (*Taeniopygia guttata*) annotation (Supplementary tables 3 to 6). Genes found in regions associated with elevation displayed enrichment for GO terms linked to development, body growth, gene expression, RNA metabolism, DNA organization, immune-system development, hematopoiesis and hemoglobin synthesis (GO:0005833: hemoglobin complex, 3 genes found over 4 in total, P=0.0031). For genes associated with each one of the parapatric lowland forms, we mostly found associations with immune response, metabolic process, hematopoiesis, reproduction, morphogenesis, and development (Supplementary Tables 4 to 6).

To identify candidates for plumage colour variation between forms, we screened the genes found in outlier regions for GO terms linked to melanin synthesis and metabolism (GO:0042438, GO:0046150, GO:0006582). *TYRP1* (*tyrosinase-related protein 1*), located on the Z chromosome, was the only gene that was systematically found associated with BNB, GHB and LBHB forms. Another gene, *WNT5A,* was found associated only with the GHB form.

Discussion

Genetic structure and genomic islands of differentiation

Our results quantify the relative importance of autosomal and sex-linked genetic variation underlying a potential case of incipient speciation in the Reunion grey white-eye. We confirm previous findings based on microsatellite data about the existence of a fine-grained population structure, with low but significant F_{ST} between localities (Bertrand et al., 2014), and a clear divergence between populations from low and high elevations (Cornuault et al., 2015; Bertrand et al., 2016). These events of divergence have likely occurred in the last 430,000 years (Warren et al. 2006), and perhaps even more recently since this divergence time estimate was based on mitochondrial data that tend to yield overestimates of population splitting times (Moore, 1995) (Moore 1995). A striking result is the clear contrast between patterns of variation in autosomes and the Z chromosome, the latter discriminating more clearly the different geographic forms and systematically displaying markers associated with plumage colour phenotypes and elevational ranges. Similar results were found independently for both pooled and individual datasets, and with two different SNP callers. These observations combined with demographic analyses suggest that while extensive gene flow has occurred between populations from high and low elevations, the Z chromosome acts as a barrier to it.

Reproductive isolation between forms may explain divergence at the Z chromosome

Reproductive isolation between nascent species can arise as incompatibilities between interacting loci accumulate in the genome, as described by the Bateson-Dobzhansky-Muller model (Dobzhansky 1934; Muller 1940, 1942; Orr 1996). Sex chromosomes are particularly likely to accumulate such incompatibilities, since the genes they harbor are not transmitted by the same rules between males and females (Seehausen et al., 2014) and are therefore prone to genetic conflicts. Indeed, sex chromosomes often harbor many genes causing disruption of fertility or lower viability in hybrids

(Storchová et al., 2004, 2010; Masly and Presgraves, 2007; Good et al., 2008), which may lead to increased divergence (Carling and Brumfield, 2008; Macholán et al., 2011; Ellegren et al., 2012), and faster emergence of postzygotic isolation (Lima, 2014). Hemizygosity of sex-chromosomes may also lead to the exposure of recessive incompatibilities in heterogametic individuals (Qvarnström and Bailey, 2009). This large effect of sex chromosomes has been a common explanation of Haldane's rule (Haldane 1922; Orr 1997; Coyne and Orr 2004), which states that in hybrids, the heterogametic sex often displays a stronger reduction in fitness, and may explain why an excess of highly differentiated regions on sex chromosomes is often observed during the early stages of speciation (e.g. Backström et al. 2010). An excess of highly differentiated regions on sex chromosomes at the early stages of speciation therefore suggests a role for intrinsic barriers to gene flow in promoting divergence (Backström et al., 2010).

In addition, premating and prezygotic isolation affect sexually dimorphic traits that are under the control of sex-linked genes (Pryke, 2010). While this type of isolation would lead to divergence across the entire genome due to the isolation of gene pools, divergence is expected to be accelerated at loci controlling traits under sexual selection, especially if hybrids and backcrosses have lower mating success (Svedin et al., 2008). Finally, effective recombination rates are lower in sex chromosomes since they recombine in only one sex (males in birds), which facilitates linkage of genes involved in pre- and post-zygotic barriers. This linkage would ultimately promote reinforcement of isolation between species or populations (Pryke, 2010). Because of these processes, gene flow at isolating sex-linked loci is expected to be impeded between diverging populations, leaving a stronger signature of differentiation than in the rest of the genome (e.g. Mořkovský et al., 2018).

Our findings are in line with these theoretical expectations, and show an excess of highly differentiated markers on the Z chromosome that display evidence for resisting gene flow, a pattern that is also consistent with Haldane's rule (e.g. Carling and Brumfield, 2008). Importantly, highly differentiated autosomal markers were mostly associated with differences in elevation ranges but Z chromosome variation was found to be associated with both elevation and plumage colour differences between lowland forms. This suggests that the divergence between lowland and highland forms is due to both polygenic adaptation to different elevational ranges and behavioural differentiation, e.g. mating preferences (Sæther et al., 2007).

We acknowledge that we have no direct evidence yet of assortative mating, character displacement, or lower hybrid fitness in contact zones between the different geographic forms of our study species, but (i) a previous study showed that Reunion grey white-eyes can perceive colour differences between forms, and that both within- and between-form differences in plumage colour can be discriminated (Cornuault et al. 2015); (ii) hybrid zones are narrow among lowland forms (Delahaie et al., 2017) and quite likely also between lowland and highland forms (Bertrand et al 2016), suggesting that hybrid phenotypes must have a lower fitness; and (iii) high gene flow within forms can erase signatures of character displacement if alleles responsible for pre-mating isolation are nearly neutral in populations that are distant from contact zones (Servedio and Noor, 2003).

Role of drift and recent selective sweeps in Z chromosome divergence

The higher differentiation observed on the Z chromosome may also be due to its reduced effective population size, which leads to faster drift and lineage sorting between forms. Such a mechanism may explain the generally higher rate of divergence observed on Z chromosomes (fast-Z effect). This effect can become even stronger in the presence of a recent bottleneck and population size change (Pool and Nielsen, 2007; Van Belleghem et al., 2018). However, we did not find any evidence for strong bottlenecks or abrupt changes in population sizes, making faster accumulation of divergent alleles in recently established populations unlikely. Moreover, another recent study on *Z. borbonicus* has also shown that there is weak support for a fast-Z effect in the clade to which this species belongs (Leroy et al., 2019). Another mechanism that might lead to stronger differentiation at Z-linked loci is female-biased dispersal, which is frequent in passerine birds (Greenwood, 1980). However, dispersal in *Z. borbonicus* is extremely reduced for both sexes (Bertrand et al. 2014), which should attenuate the contrast between autosomal and Z-linked markers. In addition, although it could lead to stronger differentiation between localities, such a mechanism alone is unlikely to explain why such a high proportion of variance is associated with forms on the Z chromosome. At last, we do not observe higher differentiation on Z markers or autosomes in males as compared to all individuals (Table 1). Lower effective recombination rates on the sex chromosome may enhance the effects of selection at sites linked to loci involved in ecological adaptation, further reducing diversity on the Z. In addition, a possible explanation of fast-Z effect in birds may result from positive selection on recessive beneficial alleles in heterogametic females (Dean et al. 2015). At last faster drift on the Z chromosome may lead

to a faster accumulation of incompatibilities (Janoušek et al., 2019), which will further increase divergence. Ultimately, drift and selection are interconnected processes that are difficult to disentangle. Unfortunately, we cannot provide with our dataset alone a detailed understanding of how processes such as linked selection, chromosomal rearrangements and barriers to gene flow can interact (Bierne et al., 2011; Ravinet et al., 2017). Future studies using whole-genome resequencing data should provide a clearer picture, by providing information on genealogies and age of both autosomal and sex-linked haplotypes.

We note however that the allele frequency spectra of lowland and highland forms do not display the expected signature of pervasive linked selection, with Tajima's *D* actually higher for Z markers than for autosomes. Our demographic analyses also show that the effective population sizes estimated from Z markers are reduced by a factor of 0.38-0.89 when compared to autosomes, the neutral expectation being 0.75. We acknowledge that polygenic adaptation to divergent environmental pressures is likely partly responsible for the observed genomic landscape of differentiation, possibly leading to localised divergence at autosomal and Z loci (see below). However, reproductive isolation seems to be at play in this system, either through premating isolation based on plumage characters, and/or post-zygotic isolation. More research about the ecology of the Reunion grey white-eye would also be useful to quantify the extent of sex-biased dispersal, assortative mating, parental imprinting, intrinsic incompatibilities, and how these factors may interact in this system (Pryke, 2010; Seehausen et al., 2014).

Autosomal divergence is mostly associated with elevation

Adaptation to local environmental conditions or natural selection against hybrids is more likely to be driven by genes scattered across the genome, assuming polygenic selection (Qvarnström and Bailey, 2009; Seehausen et al., 2014). In addition to Z-linked loci, we found many differentiated loci on autosomes, mostly in association with elevation. Genomic regions associated with elevation displayed an enrichment of genes involved in development and body growth, and included the cluster of hemoglobin subunits A, B, and Z on chromosome 14. The function of these genes is consistent with biological expectations, given the wide elevational range (from 0 to 3,000 m) occupied by white-eyes on Reunion.

Large chromosomal rearrangements are powerful drivers of differentiation, since they prevent recombination between several consecutive genes, facilitating the maintenance of divergent allele combinations between populations. A famous example of adaptive inversions facilitating the maintenance of colour (geographic) forms and species has been reported in *Heliconius* butterflies (Joron et al., 2011) or the white-throated sparrow (Tuttle et al., 2016), and these rearrangements have been predicted to take place in isolation followed by secondary contact (Feder et al., 2011). In this study, autosomal regions associated with geographic forms and elevation sometimes spanned more than 1 Mb. This suggests a possible role for large scale rearrangements and linked selection in regions of low recombination as a substrate for divergence in Reunion grey white-eyes. Our results remind at a much smaller spatial scale what has been previously observed in *Ficedula* flycatchers, with high differentiation on the Z chromosome, linked selection, and chromosomal rearrangements (Ellegren et al., 2012; Burri et al., 2015).

The low number of Z-linked markers found in this study and their higher level of population differentiation may limit the interpretation of results when comparing patterns of divergence with autosomes. Whole-genome resequencing data and refined demographic models building on the ones used in this study will be critical to precisely quantify the evolutionary dynamics of the Z chromosome compared to autosomes, and identify at a higher resolution the loci displaying strong divergence. Future studies should also focus on the variation in allele frequencies along hybrid zones and test whether loci that are more likely to be involved in local adaptation (such as immune genes or hemoglobin subunits) display changes in frequencies that are as sharp as Z-linked loci, since the latter are more likely involved in both pre- and post-zygotic isolation. Overall, our results suggest an extreme case of divergence with gene flow that can bring valuable insights into the relative order at which pre- and post- zygotic isolation mechanisms occur during speciation.

Genetics of colour

TYRP1, a well-characterized colour gene in both model species and natural populations (Nadeau et al., 2007; Backström et al., 2010; Delmore et al., 2016; Abolins-Abols et al., 2018), was the only known colour gene found in the regions associated with each of the three lowland forms. This gene

had been previously studied using a candidate gene approach in *Z. borbonicus*, but no clear association with plumage colour phenotypes was found, probably owing to the low levels of polymorphism displayed among lowland forms (Bourgeois et al., 2016). The *WNT5A* gene was found associated with the GHB form, with a brown back and a grey head. This gene is known to regulate *TYRP1* expression (Zhang et al., 2013) and is differentially expressed between black carrion and grey-coated hooded crows (Poelstra et al., 2015). However, this gene is not only involved in melanogenesis but also in cell migration and differentiation, making it a less straightforward candidate in this system.

Our pool-seq approach only covered about 10-20% of the genome, and may therefore have missed some colour loci. The likelihood is high, however, that recent selective sweeps would have been detected if such genes had been targeted by selection. For example, the locus underlying color polymorphism in the high-elevation form shows clear signs of a selective sweep reducing diversity over 500kb (see Figure 3 in Bourgeois et al. 2017), a region which is large enough to be covered by tens of RAD-seq loci. Cases of long-term balancing selection may however be harder to detect because of the short haplotypes typically found in this case due to extensive recombination. Together with previous findings (Bourgeois et al., 2012, 2017), this suggests that a large part of plumage colour variation between the geographic forms of the Reunion grey white-eye may be controlled by a set of a few loci of major effect. More detailed studies of hybrid zones between the different lowland forms may help characterize the exact association of alleles that produce a given plumage colour phenotype.

Acknowledgments

We thank Joseph Manthey, Stéphane Boissinot, Maëva Gabrielli and Thibault Leroy for insightful comments that improved the manuscript. We also thank the Reunion National Park for granting us permission to conduct fieldwork and to collect blood samples. Thomas Duval, Guillaume Gélinaud, Josselin Cornuault, Philipp Heeb, Dominique Strasberg, Ben Warren, and Juli Broggi assisted with fieldwork. Emeline Lhuillier and Olivier Bouchez assisted with the development of pooled RAD-seq. This research was carried out on the High-Performance Computing resources at New York University

Abu Dhabi and the Genotoul HPC cluster. This work was supported by Fondation pour la Recherche sur la Biodiversité (FRB), Agence Française pour le Développement (AFD), Agence Nationale de la Recherche (ANR-2006-BDIV002), Centre National de la Recherche Scientifique (CNRS) through a PEPS grant, The National Geographic Society, and the "Laboratoire d'Excellence" TULIP (ANR-10-LABX-41). The first author was supported by a MESR (Ministère de l'Enseignement Supérieur et de la Recherche) PhD scholarship during this study.

Author contributions

BM and CT initiated, coordinated and supervised the project; YB, BM and CT conceived the study and designed the experiments; BM, CT, YB, JAB and BD conducted the fieldwork; molecular data were generated by YB and HH; YB analyzed the data; and YB, BM, and CT wrote the paper with input from the other authors. All authors gave final approval for publication.

Data Accessibility

All data associated with this manuscript are published on DRYAD (VCF files,allele counts for pooled data and position of *Z. lateralis* scaffolds on zebra finch chromosomes; https://doi.org/10.5061/dryad.z34tmpg8z) and European Nucleotide Archive (BAM files for pools and fastq files for individual GBS data; accession number PRJEB36701).

References

Abolins-Abols M, Kornobis E, Ribeca P, Wakamatsu K, Peterson MP, Ketterson ED, et al. (2018) Differential gene regulation underlies variation in melanic plumage coloration in the dark-eyed

junco (*Junco hyemalis*). *Mol Ecol,* **27**, 4501–4515.

Alexander DH, Novembre J (2009) Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res*, **19**, 1655–1664.

Axelsson E, Smith NGC, Sundström H, Berlin S, Ellegren H (2004) Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and Turkey. *Mol Biol Evol,* **21**, 1538–47.

Backström N, Lindell J, Zhang Y, Palkopoulou E, Qvarnström A, Saetre G-P, et al. (2010) A high-density scan of the Z chromosome in Ficedula flycatchers reveals candidate loci for diversifying selection. *Evolution,* **64**, 3461–75.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One,* **3**, e3376.

Van Belleghem SM, Baquero M, Papa R, Salazar C, McMillan WO, Counterman BA, et al. (2018) Patterns of Z chromosome divergence among Heliconius species highlight the importance of historical demography. *Mol Ecol,* **27**, 3852–3872.

Bertrand JAM, Bourgeois YXC, Delahaie B, Duval T, García-Jiménez R, Cornuault J, et al. (2014) Extremely reduced dispersal and gene flow in an island bird. *Heredity,* **112**, 190–6.

Bertrand JAM, Delahaie B, Bourgeois YXC, Duval T, García-Jiménez R, Cornuault J, et al. (2016) The role of selection and historical factors in driving population differentiation along an elevational gradient in an island bird. *J Evol Biol,* **29**, 824–836.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis, why genome scans may fail to map local adaptation genes. *Mol Ecol,* **20**, 2044–72.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic, A flexible trimmer for Illumina sequence data. *Bioinformatics,* **30**, 2114–2120.

Bourgeois YXC, Bertrand JAM, Delahaie B, Cornuault J, Duval T, Milá B, et al. (2016) Candidate Gene Analysis Suggests Untapped Genetic Complexity in Melanin-Based Pigmentation in Birds. *J Hered,* **107**, 327–335.

Bourgeois YXC, Bertrand JAM, Thébaud C, Milá B (2012) Investigating the role of the melanocortin-1 receptor gene in an extreme case of microgeographical variation in the pattern of melanin-based plumage pigmentation. *PLoS One,* **7**, e50906.

Bourgeois YXC, Delahaie B, Gautier M, Lhuillier E, Malé P-JG, Bertrand JAM, et al. (2017) A novel locus on chromosome 1 underlies the evolution of a melanic plumage polymorphism in a wild songbird. *R Soc Open Sci*, 160805.

Bourgeois YXC, Lhuillier E, Cézard T, Bertrand J a M, Delahaie B, Cornuault J, et al. (2013) Mass production of SNP markers in a nonmodel passerine bird through RAD sequencing and contig mapping to the zebra finch genome. *Mol Ecol Resour,* **13**, 899–907.

Burri R (2017) Interpreting differentiation landscapes in the light of long-term linked selection. *Evol Lett,* **1**, 118–131.

Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Res,* **25**, 1656–1665.

Carling MD, Brumfield RT (2008) Haldane's rule in an avian system, using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the Passerina bunting hybrid zone. *Evolution,* **62**, 2600–15.

Cornetti L, Valente LM, Dunning LT, Quan X, Black RA, Hébert OH, et al. (2015) The genome of the 'great speciator' provides insights into bird diversification. *Genome Biol Evol,* **7**, 2680–2691.

Cornuault J, Delahaie B, Bertrand JAM, Bourgeois YXC, Mila B, Heeb P, et al. (2015) Morphological and plumage colour variation in the Réunion grey white-eye (Aves, Zosterops borbonicus), Assessing the role of selection. *Biol J Linn Soc,* **114**, 459–473.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol,* **23**, 3133–3157.

Csilléry K, François O, Blum MGB (2012) abc, an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol,* **3**, 475–479.

Dean, R., Harrison, P. W., Wright, A. E., Zimmer, F., & Mank, J. E. (2015). Positive Selection

Underlies Faster-Z Evolution of Gene Expression in Birds. *Molecular biology and evolution*, **32**, 2646–2656.

Delahaie B, Cornuault J, Masson C, Bertrand JAM, Bourgeois YXC, Milá B, et al. (2017) Narrow hybrid zones in spite of very low population differentiation in neutral markers in an island bird species complex. *J Evol Biol,* **30**, 2132–2145.

Delmore KE, Hübner S, Kane NC, Schuster R, Andrew RL, Câmara F, et al. (2015) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Mol Ecol,* **24**, 1873–1888.

Delmore KE, Toews DPL, Germain RR, Owens GL, Irwin DE (2016) The Genetics of Seasonal Migration and Plumage Color. *Curr Biol,* **26**, 2167–2173.

Derjusheva S, Kurganova A, Habermann F, Gaginskaya E (2004) High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds. *Chromosome Res,* **12**, 715–23.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, et al. (2012) The genomic landscape of species divergence in Ficedula flycatchers. *Nature,* **491**, 756–60.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One,* **6**, e19379.

Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M (2013) Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet,* **9**.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5, a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour,* **10**, 564–7.

Feder JL, Gejji R, Powell THQ, Nosil P (2011) Adaptive chromosomal divergence driven by mixed geographic mode of evolution. *Evolution,* **65**, 2157–2170.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv Prepr arXiv12073907*, 9.

Gautier M (2015) Genome-Wide Scan for Adaptive Divergence and Association with Population-

Specific Covariates. *Genetics,* **201**, 1555–1579.

Gavrilets S (2014) Models of speciation, Where are we now? *J Hered,* **105**, 743–755.

Gill FB (1973) Intra-island variation in the Mascarene White-eye Zosterops borbonica. *Ornithol Monogr* **12**.

Gill F, Donsker D (2015) IOC World Bird List v5.1. *Int Ornithol Union Comm Nomencl*.

Good JM, Dean MD, Nachman MW (2008) A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics,* **179**, 2213–2228.

Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. *Anim Behav,* **28**, 1140–1162.

Griffiths R, Double M, Orr K, Dawson RJG (1998) A DNA test to sex most birds. *Mol Ecol,* **7**, 1071–1075.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet,* **5**.

Harris RS (2007) *Improved pairwise alignment of genomic DNA.*

Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R (2018) Measuring genetic differentiation from pool-seq data. *Genetics,* **210**, 315–330.

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods,* **12**, 115–121.

Janoušek V, Fischerová J, Mořkovský L, Reif J, Antczak M, Albrecht T, Reifová R.. (2019). Postcopulatory sexual selection reduces Z-linked genetic variation and might contribute to the large Z effect in passerine birds. *Heredit,y* **122**,622-635.

Jeffries DL, Copp GH, Handley LL, Håkan Olsén K, Sayer CD, Hänfling B (2016) Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, Carassius carassius, L. *Mol Ecol,* **25**, 2997–3018.

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, et al. (2011) Chromosomal

rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature,* **477**, 203–6.

Kofler R, Pandey RV, Schlötterer C (2011) PoPoolation2, identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics,* **27**, 3435–6.

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD, Analysis of Next Generation Sequencing Data. *BMC Bioinformatics,* **15**, 356.

Kozma R, Melsted P, Magnússon KP, Höglund J (2016) Looking into the past - The reaction of three grouse species to climate change over the last million years using whole genome sequences. *Mol Ecol,* **25**, 570–580.

Leroy T, Anselmetti Y, Tilak M-K, Berard S, Csukonyi L, Gabrielli M, et al. (2019) A bird's white-eye view on avian sex chromosome evolution. *bioRxiv*, 505610.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* **25**, 1754–60.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* **25**, 2078–9.

Lima TG (2014) Higher levels of sex chromosome heteromorphism are associated with markedly stronger reproductive isolation. *Nat Commun,* **5**, 4743.

Macholán M, Baird SJE, Dufková P, Munclinger P, Bímová BV, Piálek J (2011) Assessing multilocus introgression patterns, A case study on the mouse x chromosome in central europe. *Evolution,* **65**, 1428–1446.

Masly JP, Presgraves DC (2007) High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS Biol,* **5**, 1890–1898.

Milá B, Warren BH, Heeb P, Thébaud C (2010) The geographic scale of diversification on islands, genetic and morphological divergence at a very small spatial scale in the Mascarene grey white-eye (Aves, *Zosterops borbonicus*). *BMC Evol Biol,* **10**, 158.

Moore WS (1995) Inferring Phylogenies from mtDNA Variation, Mitochondrial-Gene Trees Versus

Nuclear-Gene Trees. *Evolution* **49**, 718–726.

Mořkovský L, Janoušek V, Reif J, Rídl J, Pačes J, Choleva L, et al. (2018) Genomic islands of differentiation in two songbird species reveal candidate genes for hybrid female sterility. *Mol Ecol,* **27**, 949–958.

Nadeau NJ, Mundy NI, Gourichon D, Minvielle F (2007) Association of a single-nucleotide substitution in TYRP1 with roux in Japanese quail (Coturnix japonica). *Anim Genet,* **38**, 609–13.

Nazareno AG, Bemmels JB, Dick CW, Lohmann LG (2017) Minimum sample sizes for population genomics, an empirical study from an Amazonian plant species. *Mol Ecol Resour,* **17**, 1136–1147.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends Ecol Evol,* **24**, 145–56.

Otto SP, Day T (2007) *A biologist's guide to mathematical modeling in ecology and evolution, volume 13.* Princeton University Press.

Pala I, Naurin S, Stervander M, Hasselquist D, Bensch S, Hansson B (2012) Evidence of a neo-sex chromosome in birds. *Heredity,* **108**, 264–272.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet,* **2**, 2074–2093.

Poelstra JW, Vijay N, Hoeppner MP, Wolf JBW (2015) Transcriptomics of colour patterning and coloration shifts in crows. *Mol Ecol,* **24**, 4617–4628.

Pool JE, Nielsen R (2007) Population size changes reshape genomic patterns of diversity. *Evolution,* **61**, 3001–3006.

Pryke SR (2010) Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. *Evolution,* **64**, 1301–10.

Quinlan AR, Hall IM (2010) BEDTools, a flexible suite of utilities for comparing genomic features. *Bioinformatics,* **26**, 841–842.

Qvarnström A, Bailey RI (2009) Speciation through evolution of sex-linked genes. *Heredity,* **102**, 4–

15.

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, et al. (2017) Interpreting the genomic landscape of speciation, finding barriers to gene flow. *J Evol Biol,* **30**, 1450–1477.

Robinson JD, Coffman AJ, Hickerson MJ, Gutenkunst RN (2014) Sampling strategies for frequency spectrum-based population genomic inference. *BMC Evol Biol*, **14**, 1–16.

Sæther SA, Sætre GP, Borge T, Wiley C, Svedin N, Andersson G, et al. (2007) Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science,* **318**, 95–97.

Sætre GP, Sæther SA (2010) Ecology and genetics of speciation in Ficedula flycatchers. *Mol Ecol,* **19**, 1091–1106.

Safran RJ, Scordato ESC, Symes LB, Rodríguez RL, Mendelson TC (2013) Contributions of natural and sexual selection to the evolution of premating reproductive isolation, A research agenda. *Trends Ecol Evol,* **28**, 643–650.

Schwartz S, Kent W, Smit A (2003) Human–mouse alignments with BLASTZ. *Genome Res*, **13**, 103–107.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, et al. (2014) Genomics and the origin of species. *Nat Rev Genet,* **15**, 176–192.

Servedio MR, Noor MAF (2003). The Role of Reinforcement in Speciation, Theory and Data. *Annu Rev Ecol Evol Syst,* **34**, 339–364.

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. *Can J Zool*, **69**, 82–90.

Storchová R, Gregorová S, Buckiová D, Kyselová V, Divina P, Forejt J (2004) Genetic analysis of X-linked hybrid sterility in the house mouse. *Mamm Genome,* **15**, 515–524.

Storchová R, Reif J, Nachman MW (2010) Female heterogamety and speciation, Reduced introgression of the z chromosome between two species of nightingales. *Evolution,* **64**, 456–471.

Svedin N, Wiley C, Veen T, Gustafsson L, Qvarnstrom A (2008) Natural and sexual selection against

hybrid flycatchers. *Proc R Soc B Biol Sci,* **275**, 735–744.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics,* **123**, 585–95.

Takezaki N, Nei M, Tamura K (2010) POPTREE2, Software for constructing population trees from allele frequency data and computing other population statistics with windows interface. *Mol Biol Evol,* **27**, 747–752.

Tuttle EM, Bergland AO, Korody ML, Brewer MS, Newhouse DJ, Minx P, et al. (2016) Divergence and functional degradation of a sex chromosome-like supergene. *Curr Biol,* **26**, 344–350.

Warren BH, Bermingham E, Prys-Jones RP, Thébaud C (2006) Immigration, species radiation and extinction in a highly diverse songbird lineage, white-eyes on Indian Ocean islands. *Mol Ecol,* **15**, 3769–86.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, et al. (2010) The genome of a songbird. *Nature,* **464**, 757–62.

Willing EM, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by fst do not necessarily require large sample sizes when using many snp markers. *PLoS One,* **7**, 1–7.

Wolf JBW, Ellegren H (2016) Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet,* **18**, 87–100.

Wu CI (2001) The genic view of the process of speciation. *J Evol Biol,* **14**, 851–865.

Zhang J, Li Y, Wu Y, Yang T, Yang K, Wang R, et al. (2013) Wnt5a inhibits the proliferation and melanogenesis of melanocytes. *Int J Med Sci,* **10**, 699–706.

Figures

Figure 1. Map of Reunion showing *Zosterops* localities sampled in this study and a description of

population structure using Principal Components Analysis (PCA) on autosomal GBS data. A: For each locality, the sample size for the individual GBS dataset is followed by the sample size for each pool. In the three localities found at high elevation, where populations are polymorphic, the size of pools is given for grey + brown individuals. Distribution limits between the different geographic forms are indicated by dashed lines. B: PCA results including *Zosterops mauritianus* (left panel) or not (right panels). Points corresponding to high elevation individuals were removed on the last three panels for the sake of clarity.

Figure 2. A: Co-ancestry coefficients obtained from ADMIXTURE for K values ranging from 2 to 7. Population codes as in Figure 1. Separate analyses were run on autosomal and Z markers. Bold and thin vertical lines indicate limits between forms and populations, respectively. B: POPTREE2 analysis obtained from 20,000 markers randomly sampled from pooled data. Branch support was obtained from 1,000 bootstraps.

Figure 3. Summary of the four demographic models and their parameters evaluated using GBS data .

Figure 4. Manhattan plots of X^TX and Bayesian empirical *P*-values (*eBPis*) for association with elevation and geographic form. Note that X^TX are not directly comparable for autosomes and Z chromosomes as they were analyzed separately and display different demographic histories (see Figure 3).

Tables

Table 1: Pairwise AMOVAs comparing geographic forms within Reunion grey white-eye and between Reunion and Mauritius grey white-eyes. All values are significant (1000 permutations). To properly compare patterns observed on sex-linked and autosomal markers, results for autosomal markers using only males are also shown.

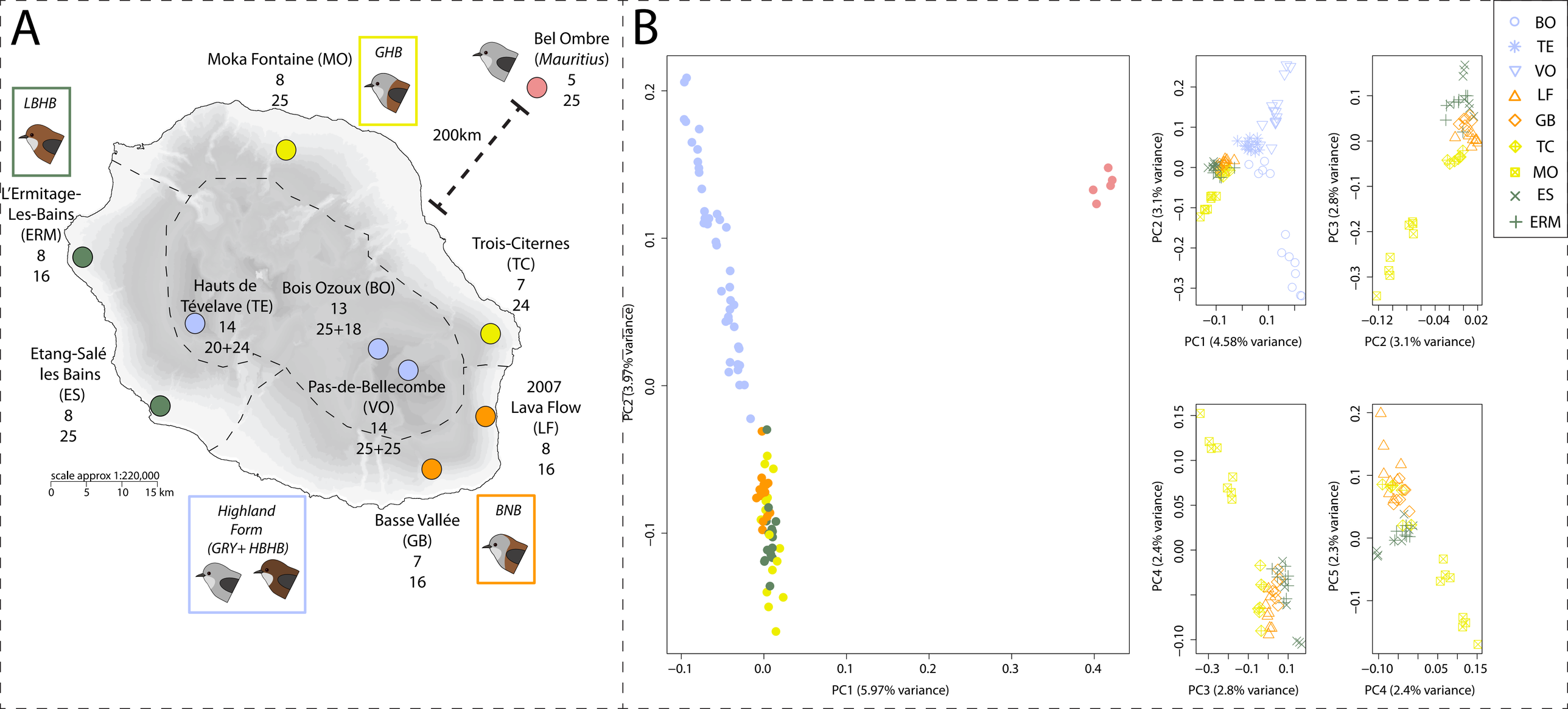
Comparison	LBHB v. BNB		GHB v. LBHB		GHB v. BNB		High v. Low		Reunion v. Mauritius	
Autosomes (wo 4A sex-linked markers)	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained
Among groups (*Fct*)	0.005	0.48	0.010	1.03	0.006	0.59	0.016	1.57	0.173	17.31
Among populations within groups (Fsc)	0.033	3.31	0.042	4.11	0.035	3.49	0.040	3.97	0.049	4.09
Within populations (Fst)	0.038	96.21	0.051	94.86	0.041	95.92	0.052	94.46	0.214	78.61
Z-linked markers	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained
Among groups (Fct)	0.068	6.82	0.108	10.84	0.034	3.45	0.126	12.59	0.206	20.61
Among populations within groups (Fsc)	0.038	3.56	0.093	8.28	0.095	9.13	0.123	10.78	0.191	15.17
Within populations (Fst)	0.104	89.62	0.191	80.88	0.095	87.42	0.234	76.63	0.358	64.21
Autosomes (wo 4A sex-linked markers, Males only)	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained
Among groups (Fct)	0.007	0.69	0.009	0.93	0.005	0.53	0.019	1.86	0.166	16.6
Among populations within groups (Fsc)	0.013	1.32	0.029	2.87	0.027	2.72	0.040	3.92	0.051	4.25
Within populations (Fst)	0.02	97.98	0.038	96.2	0.032	96.75	0.058	94.22	0.209	79.15
4A-1-10Mb (Males only)	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained	F-statistics	% of variance explained
Among groups (Fct)	0.044	4.44	0.069	6.92	0.048	4.81	0.080	7.95	0.370	28.58
Among populations within groups (Fsc)	0.050	4.80	0.059	5.47	0.039	3.68	0.075	6.89	0.118	8.44
Within populations (Fst)	0.044	90.76	0.124	87.61	0.085	91.51	0.148	85.16	0.286	62.98

Table 2. Likelihoods of the four demographic models compared with fastsimcoal2.6, and their AIC scores. Note that we report likelihoods in natural logs, while likelihoods produced by fastsimcoal are expressed as log10.

Markers	Model	Log(Likelihood)	# parameters	AIC	Δ AIC
Autosomes	D	-36,333.59	11	72,689.19	0.00
	C	-36,372.41	8	72,760.81	71.62
	B	-36,453.23	6	72,918.45	229.27
	A	-37,509.72	7	75,033.44	2,344.25
Z-linked	D	-1611.58	11	3,245.17	0.00
	C	-1628.72	8	3,273.44	28.28
	B	-1662.54	6	3,337.09	91.92
	A	-1957.25	7	3,928.51	683.34

Table 3: Parameter estimates for a model with migration rates using SFS from ANGSD. Population sizes are haploid sizes (with N representing the number of diploid individuals).

Markers	Parameter	$2N_{Mauritius}$	$2N_{Low}$	$2N_{High}$	$2N_{anc\_Reu}$	$2N_{anc\_Reu+Mau}$	T_{split}	T_{change}	$m_{Low\rightarrow High}$ (recent)	$m_{High\rightarrow Low}$ (recent)	$m_{Low\rightarrow High}$ (ancient)	$m_{High\rightarrow Low}$ (ancient)
Autosomal	Best estimate	1081493	2957776	606586	1266228	1022327	400228	79188	1.63E-05	2.03E-06	2.00E-07	3.11E-06
Autosomal	2.5% lower bound	906204	2303253	480897	596984	916712	374624	62305	1.54E-05	2.76E-07	1.40E-08	2.36E-06
Autosomal	97.5% upper bound	1158440	2957776	686191	1332774	1140005	423302	204547	2.17E-05	3.60E-06	3.82E-06	5.64E-06
Z-linked	Best estimate	640280	1595330	228792	996150	905370	404812	133746	6.66E-06	3.10E-07	1.92E-08	2.28E-09
Z-linked	2.5% lower bound	490028	1195174	140680	596824	565930	348043	66246	4.34E-06	3.76E-08	1.47E-09	6.91E-10
Z-linked	97.5% upper bound	878312	1807538	338704	1175232	1194501	419568	166238	1.05E-05	1.14E-06	2.36E-06	2.73E-07


A
Moka Fontaine (MO)
8
25
GHB
Bel Ombre
(Mauritius)
5
25
200km
LBHB
L'Ermitage-
Les-Bains
(ERM)
8
16
Trois-Citernes
(TC)
7
24
Hauts de
Tévelave (TE)
14
20+24
Bois Ozoux (BO)
13
25+18
Etang-Salé
les Bains
(ES)
8
25
Pas-de-Bellecombe
(VO)
14
25+25
2007
Lava Flow
(LF)
8
16
scale approx 1:220,000
0 5 10 15 km
Highland
Form
(GRY+ HBHB)
Basse Vallée
(GB)
7
16
BNB
B
PC2 (3.97% variance)
PC1 (5.97% variance)
PC2 (3.1% variance)
PC1 (4.58% variance)
PC3 (2.8% variance)
PC2 (3.1% variance)
PC4 (2.4% variance)
PC3 (2.8% variance)
PC5 (2.3% variance)
PC4 (2.4% variance)
BO
TE
VO
LF
GB
TC
MO
ES
ERM

A

Autosomal markers

Z markers

BO VO TE LF GB MO TC ERM ES MAU

K=2

K=3

K=4

K=5

K=6

K=7

Coancestry coefficients

0.0 0.4 0.8

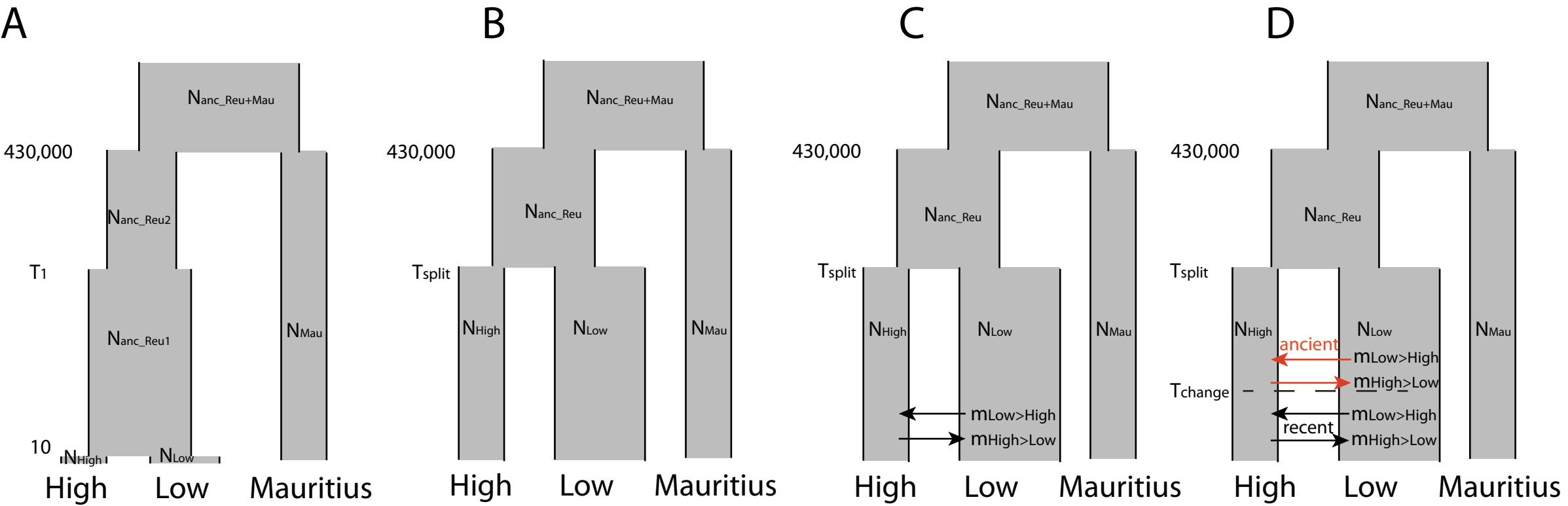
B

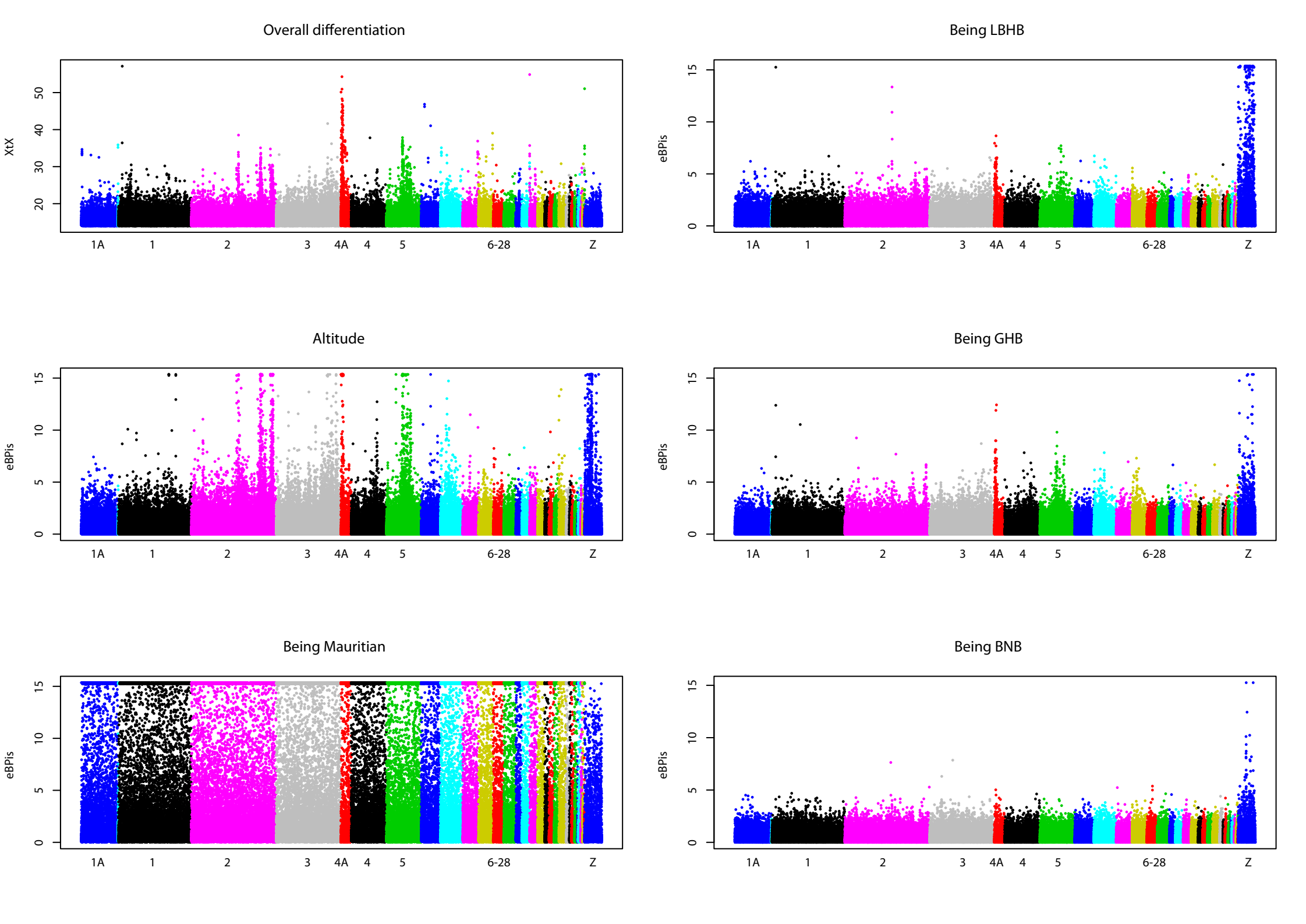
Autosomal markers (Pools)

MO
BO-HBHB
BO-GRY
VO-HBHB
VO-GRY
TE-HBHB
TE-GRY
GB
LF
TC
ERM
ES
MAU
1
0.57
0.764
0.632
0.987
0.801
0.03

Z markers (Pools)

BO-HBHB
BO-GRY
VO-HBHB
VO-GRY
TE-GRY
TE-HBHB
MO
TC
GB
LF
ERM
ES
MAU
1
0.52
0.98
0.04


A
Nanc_Reu+Mau
430,000
Nanc_Reu2
T1
Nanc_Reu1
NMau
10
NHigh
NLow
High
Low
Mauritius
B
Nanc_Reu+Mau
430,000
Nanc_Reu
Tsplit
NHigh
NLow
NMau
High
Low
Mauritius
C
Nanc_Reu+Mau
430,000
Nanc_Reu
Tsplit
NHigh
NLow
NMau
mLow>High
mHigh>Low
High
Low
Mauritius
D
Nanc_Reu+Mau
430,000
Nanc_Reu
Tsplit
NHigh
NLow
NMau
ancient
mLow>High
Tchange
mHigh>Low
mLow>High
recent
mHigh>Low
High
Low
Mauritius


Overall differentiation
XtX
20
30
40
50
1A
1
2
3
4A
4
5
6-28
Z
Being LBHB
eBPIs
0
5
10
15
Altitude
Being GHB
Being Mauritian
Being BNB